

Positioning Errors of Pencil-beam Interferometers for Long Trace Profilers

Valeriy V. Yashchuk*

Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, USA 94720

ABSTRACT

We analyze the random noise and the systematic errors of the positioning of the interference patterns in the long trace profilers (LTP). The analysis, based on linear regression methods, allows the estimation of the contributions to the positioning error of a number of effects, including non-uniformity of the detector photo-response and pixel pitch, read-out and dark signal noise, ADC resolution, as well as signal shot noise. The dependence of the contributions on pixel size and on total number of pixels involved in positioning is derived analytically. The analysis, when applied to the LTP II available at the ALS optical metrology laboratory, has shown that the main source for the random positioning error of the interference pattern is the read-out noise estimated to be $\sim 0.2 \mu\text{rad}$. The photo-diode-array photo-response and pixel pitch non-uniformity determine the magnitude of the systematic positioning error and are found to be $\sim 0.3 \mu\text{rad}$ for each of the effects. Recommendations for an optimal fitting strategy, detector selection and calibration are provided. Following these recommendations will allow the reduction of the error of LTP interference pattern positioning to a level adequate for the slope measurement with $0.1\text{-}\mu\text{rad}$ accuracy.

Keywords: pencil-beam interferometer, long trace profiler, LTP, beam positioning, optical metrology, error reduction

1. INTRODUCTION

The long trace profiler (LTP) is the basic metrology tool for high accuracy testing the figure of X-ray optics with slope variations on the order of one μrad rms.¹⁻⁶ The LTP optical schematic is a realization of the pencil-beam interferometer. In the interferometer, two parallel light beams, possessing a phase difference, are made to interfere at the focus of the Fourier transform lens. The resulted interference fringe pattern (Fig.1) recorded with a position-sensitive detector placed at the focus has two strongly marked peaks with a minimum between the peaks. The position of the minimum is a measure of slope of the mirror surface at the place of the beam reflection. The LTP records the local slope profile of a surface by measuring the reflection angle of a laser sample beam as the beam is transported across the surface by an air bearing carriage. Due to the translation of the optical sensor, the LTP has a unique capability for surface figure metrology of very long, meter size, mirrors.⁷⁻⁹ However, non-idealities of the translation mechanism as well as non-idealities of the optical elements bring forth a number of systematic effects, which often dominate over the random noise of the LTP measurement.¹⁰⁻¹² Some of the systematic effects have been analyzed and the methods for their suppression have been developed.¹³⁻¹⁷

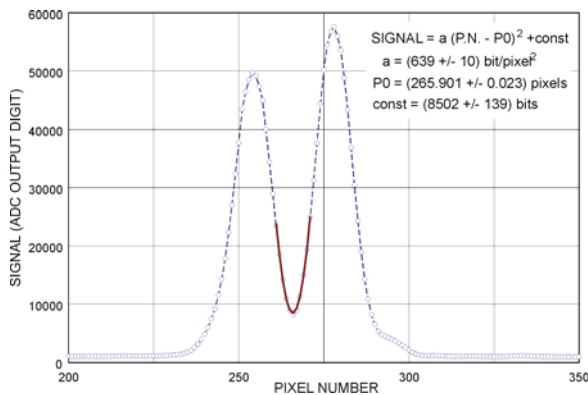


Figure 1: A characteristic interference pattern recorded with the LTP II at the ALS Optical Metrology Laboratory (OML). The LTP detector is based on a linear photo-diode array, consisting of 1024 pixels with a pitch of $25 \mu\text{m}$ and sagittal width of 2.5 mm . With such a detector, number of points usually used for positioning is about 11. The focal distance of the LTP II is $F=1.250 \text{ m}$; a coefficient of detector-position-to-slope-angle transformation is $1/(2F)$, where F is given in *meters*, and slope angle and beam position on the detector are measured in μrad and μm , respectively. The solid line shows the best-fit second order polynomial fitted over data for 11 point (pixels). The parameters of the fit are presented in the figure.

*vvyashchuk@lbl.gov; phone +1 510-495-2592; fax +1 510-486-7696

In the present work, the random noise and the systematic errors of the positioning of the interference patterns in the LTP are analyzed. The analysis is based on linear regression methods¹⁸⁻²¹ briefly reviewed in Sec. 2. The estimations of the contributions to the positioning error of a number of effects, including non-uniformity of the detector photo-response and pixel pitch, read-out and dark signal noise, ADC resolution, as well as signal shot noise are given in Sec. 3. In Sec. 4 recommendations for an optimal fitting strategy, detector selection and calibration are provided. Following these recommendations will allow one to reduce the positioning error of LTP to a level adequate for the slope measurement with 0.1- μ rad accuracy.

2. BRIEF REVIEW OF THE LINEAR REGRESSION METHOD

Taking into account a limited accessibility of Ref.¹⁹ chosen by the author as a preferable reference due to its compact format of comprehensive contents, we start from a condensed narration of the mathematical method aiming to provide the reader with a basis, self-sufficient for other applications. A more detailed discussion of the method with successive proofs of formulae presented here can be found in more fundamental books.^{18,20,21}

Let us consider two variables x, y such that for any given value of x there is a conditional distribution of y . The term regression is used in statistics to define the functional dependence of a mean value (expectation) of one variable on other variables and parameters:

$$E(y | x) = \eta(x, \theta), \quad (1)$$

where θ denotes a set of unknown parameters, which completely determine the function $\eta(x, \theta)$. The classical approach to finding the parameters θ is to fit randomly selected observations of the bivariate distribution (x, y) to the function $\eta(x, \theta)$ by minimizing the sum of the squared deviations in the y -direction, the Method of Least Squares. The Linear Regression Method provides a way for calculation of the parameters θ in the case when the function $\eta(x, \theta)$ can be expanded into a linear (over the parameters θ) combination of some known functions:

$$\eta(x, \theta) = \theta_0 f_0(x) + \theta_1 f_1(x) + \dots + \theta_r f_r(x). \quad (2)$$

Keeping in mind the fitting of the LTP interference pattern, we limit ourselves to the Taylor expansion with the functions $f_\rho(x) = x^\rho$:

$$\eta(x, \theta) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_r x^r. \quad (3)$$

In general, the observations y_i can be written as

$$y_i = \eta(x_i, \theta) + \varepsilon_i, \quad (4)$$

where ε is the error variable, which determines the spread of observations y_i around expectations $E(y_i | x_i)$:

$$\varepsilon_i = y_i - \theta_0 - \theta_1 x_i - \theta_2 x_i^2 - \dots - \theta_r x_i^r, \quad i = 1, 2, \dots, n. \quad (5)$$

Assuming that dispersion functions of the ε_i are identical with equal variance σ^2 , the best-fitting regression function (3) corresponds to the minimum of the sum of squared deviations of the n observations:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i - \theta_2 x_i^2 - \dots - \theta_r x_i^r)^2, \quad (6)$$

while the parameters θ_ρ are varied

$$\frac{\partial S}{\partial \theta_\rho} = -2 \sum_{i=1}^n x_i^\rho (y_i - \theta_0 - \theta_1 x_i - \theta_2 x_i^2 - \dots - \theta_r x_i^r) = 0, \quad \rho = 0, 1, \dots, r. \quad (7)$$

The system of equations (7) can be transformed to the *normal* equations

$$\theta_0 \sum_{i=1}^n x_i^\rho + \theta_1 \sum_{i=1}^n x_i^{\rho+1} + \dots + \theta_r \sum_{i=1}^n x_i^{\rho+r} = \sum_{i=1}^n x_i^\rho y_i, \quad \rho = 0, 1, \dots, r. \quad (8)$$

The solution of the system (8) can be simplified if one uses matrix approach. First, introduce a $n \times (r+1)$ matrix termed the *regression* matrix,

$$\hat{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^r \\ 1 & x_2 & x_2^2 & \dots & x_2^r \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^r \end{bmatrix} \quad (9)$$

and a $1 \times (r+1)$ vector of parameters,

$$\hat{\theta}' = (\theta_0, \theta_1, \theta_2, \dots, \theta_r), \quad (10)$$

where the prime denotes a transposed matrix. Note that the matrix approach is also applicable for the more general case, when, e.g. the first line in (9) looks like,

$$[f_0(x_1), f_1(x_1), \dots, f_r(x_1)]. \quad (11)$$

In matrix form, equations (5) and (6) can be rewritten as

$$\hat{\varepsilon} = \hat{y} - \hat{A} \hat{\theta}, \quad (5')$$

$$S = \hat{\varepsilon}' \hat{\varepsilon} = (\hat{y} - \hat{A} \hat{\theta})' (\hat{y} - \hat{A} \hat{\theta}) = \hat{y}' \hat{y} - 2 \hat{\theta}' \hat{A}' \hat{y} + \hat{\theta}' \hat{A}' \hat{A} \hat{\theta}. \quad (6')$$

By differentiating, one can get the system of equations equivalent to Eqs. (7) and (8)

$$-2 \hat{A}' \hat{y} + 2 \hat{A}' \hat{A} \hat{\theta} = 0; \quad (7')$$

$$(\hat{A}' \hat{A}) \hat{\theta} = \hat{A}' \hat{y}. \quad (8')$$

If the matrix $\hat{A}' \hat{A}$ is a full rank matrix, the system solution is

$$\hat{\theta}^* = (\hat{A}' \hat{A})^{-1} \hat{A}' \hat{y}, \quad (12)$$

A singularity can appear in the case when the estimation is performed for an excessive number of parameters. In other words, excessiveness is observed if any lines of matrix \hat{A} are linearly dependent. It can be shown that the estimate (12) gives an unbiased estimate of the parameters $\hat{\theta}$, meaning the expectation of $\hat{\theta}^*$ is $\hat{\theta}$. Moreover, the estimate $\hat{\theta}^*$ is the most accurate among all possible unbiased estimates. In the case of independent observations y_i with equal variance σ^2 ,

$$D \hat{y} = \sigma^2 \hat{I}, \quad (13)$$

the dispersions for parameters θ can be found with a simple relation

$$D \hat{\theta}^* = ((\hat{A}' \hat{A})^{-1} \hat{A}') D \hat{y} ((\hat{A}' \hat{A})^{-1} \hat{A}')' = \sigma^2 (\hat{A}' \hat{A})^{-1}. \quad (14)$$

If the value of σ^2 is unknown, its unbiased estimate σ_e^2 can be found from the sum of squares of the differences between the extrapolating function (3) and the observations y_i and the difference between the number of observations n and the number of parameters $(r+1)$:

$$R = S_{\min} = (\hat{y} - \hat{A} \hat{\theta}^*)' (\hat{y} - \hat{A} \hat{\theta}^*) = \hat{y}' \hat{y} - \hat{y}' \hat{A} \hat{\theta}^*, \quad (15)$$

$$\sigma_e^2 = \frac{R}{n-r-1}. \quad (16)$$

The problem can be easily generalized from the case of equal dispersions given by (13) to the case of observables with dispersions determined by a positive definite matrix \hat{W} , known exactly:

$$D\hat{y} = \sigma^2 \hat{W}. \quad (17)$$

An elementary example of this kind is the dispersion of ε_i weighted with a parameter ω_i so that

$$D(\varepsilon_i) = \sigma^2 / \omega_i \quad (18)$$

that require to minimize the value of [compare with (6)]

$$S = \sum_{i=1}^n \omega_i \varepsilon_i^2. \quad (19)$$

In this case, the system of normal equations (8) will look like (assuming $\sum \omega_i = n$)

$$\theta_0 \sum_{i=1}^n \omega_i x_i^\rho + \theta_1 \sum_{i=1}^n \omega_i x_i^{\rho+1} + \dots + \theta_r \sum_{i=1}^n \omega_i x_i^{\rho+r} = \sum_{i=1}^n \omega_i x_i^\rho y_i, \quad \rho = 0, 1, \dots, r. \quad (20)$$

In order to write (18) in matrix form, we construct a diagonal matrix

$$\hat{V} = \begin{bmatrix} (\sqrt{\omega_1})^{-1} & 0 & 0 \dots & 0 \\ 0 & (\sqrt{\omega_2})^{-1} & 0 \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \dots & (\sqrt{\omega_n})^{-1} \end{bmatrix}, \quad (20)$$

which relates to the weight matrix \hat{W} via equation

$$\hat{W} = \hat{V} \hat{V}'; \quad \hat{V} = \hat{V}'. \quad (21)$$

We construct also the ‘weighted’ *regression* matrix [compare with (11)] and ‘weighted’ matrix of observations

$$\tilde{A} = \hat{V}^{-1} \hat{A}, \quad (22)$$

$$\tilde{y} = \hat{V}^{-1} \hat{y}. \quad (23)$$

Using these notations, the system (19) can be presented in matrix form analogous to (6')

$$S = \tilde{\varepsilon}' \tilde{\varepsilon} = (\tilde{y} - \tilde{A} \hat{\theta})' (\tilde{y} - \tilde{A} \hat{\theta}) = \tilde{y}' \tilde{y} - 2 \hat{\theta}' \tilde{A}' \tilde{y} + \hat{\theta}' \tilde{A}' \tilde{A} \hat{\theta}. \quad (6'')$$

And therefore, the equations for finding parameters θ will look similar to (12) as well as the equations for the dispersions of the parameters θ will look similar to (14)

$$\hat{\theta}^* = (\tilde{A}' \tilde{A})^{-1} \tilde{A}' \tilde{y} \hat{\theta}^* = (\hat{A}' \hat{W}^{-1} \hat{A})^{-1} \hat{A}' \hat{W}^{-1} \hat{y}, \quad (24)$$

$$D(\hat{\theta}) = \sigma^2 (\tilde{A}' \tilde{A})^{-1} = \sigma^2 (\hat{A}' \hat{W}^{-1} \hat{A})^{-1}. \quad (25)$$

Finally, one can modify the equations (15) and (16) to get an unbiased estimate σ^2 for the ‘weighted’ observations

$$\sigma_e^2 = \frac{R}{n-r-1} = \frac{(\hat{y} - \hat{A} \hat{\theta}^*)' \hat{W}^{-1} (\hat{y} - \hat{A} \hat{\theta}^*)}{n-r-1}. \quad (26)$$

Note that expressions (24)-(26) are valid for more general case of dispersion matrix given by (17).

3. APPLICATION OF REGRESSION ANALYSIS TO LTP FITTING

The mathematical method reviewed in Sec. 2 provides a convenient tool for finding the parameters and their errors of the best-fit parabola while fitting the interference patterns in order to find the position of the central minimum. Consider fitting an interference pattern similar to one shown in Fig. 1. The fitting function is the second order polynomial:

$$\eta(x, \theta) = \theta_0 + \theta_1 x + \theta_2 x^2, \quad (3.1)$$

where x is a coordinate variable normalized to a pixel size. The corresponding regression matrix (9) is $n \times 3$ matrix with n - number of the point used for position fitting and the vector of parameters (10) is a 1×3 rank vector. The fitting is performed to find the position of the central minimum of the interference pattern, which can be estimated via the regression parameters, and to estimate the mean square error of the position via variations of the parameters:

$$x_{\min} = -\frac{\theta_1^*}{2\theta_2^*}, \quad (3.2)$$

$$s^2(x_{\min}) = x_{\min}^2 \left\{ \frac{s^2(\theta_1^*)}{(\theta_1^*)^2} + \frac{s^2(\theta_2^*)}{(\theta_2^*)^2} \right\}, \quad (3.3)$$

where $s^2(\theta_1^*)$ and $s^2(\theta_2^*)$ are the estimations for corresponding dispersions, equal to the diagonal elements of the error matrix (25). Note that equation (3.3) assumes independence of the parameters, i.e. the corresponding non-diagonal correlation coefficients are equal to zero.

Below, we consider the positioning errors related to different sources of random noise and systematic errors while measuring the light intensity with the LTP photo-diode array (PDA).

3.1. Positioning errors due to intensity independent noise: photo-detector dark signal and read-out error

In the case of random noise, independent of the value of the measured light intensity \hat{y} , the variations of \hat{y} are accounted for with a diagonal dispersion matrix (13) of rank $n \times n$ with parameter σ^2 , $D\hat{y} = \sigma^2 \hat{I}$. For simplicity, choose the set of variables (x_i) as a centered set of $n = 2(m+1)$ pixels

$$(x_i) = (-m, -(m-1), \dots, -i, \dots, -1, 0, 1, \dots, i, \dots, (m-1), m). \quad (3.1.1)$$

With such a choice, the elements of matrix $(\hat{A}' \hat{A})$ depend only on the value of m

$$(\hat{A}' \hat{A}) = \begin{bmatrix} X0 & 0 & X2 \\ 0 & X2 & 0 \\ X2 & 0 & X4 \end{bmatrix}, \quad \text{where} \quad (3.1.2)$$

$$X0 = \sum_{i=-m}^m x_i^0 = (2m+1); X2 = \sum_{i=-m}^m x_i^2 = \frac{1}{3} m(m+1)(2m+1); X4 = \sum_{i=-m}^m x_i^4 = \frac{1}{15} m(m+1)(2m+1)(3m^2 + 3m - 1). \quad (3.1.3)$$

The inverted matrix $(\hat{A}' \hat{A})^{-1}$ expressed in the terms of Eqs. (3.1.3) is:

$$(\hat{A}' \hat{A})^{-1} = \frac{1}{X_0 X_2 X_4 - X_2^3} \begin{bmatrix} X_2 X_4 & 0 & -X_2^2 \\ 0 & X_0 X_4 - X_2^2 & 0 \\ -X_2^2 & 0 & X_0 X_2 \end{bmatrix}. \quad (3.1.4)$$

The estimated dispersions of the regression parameters, equal to the diagonal elements of (14) after substitution (3.1.3) and (3.1.4), are

$$\begin{aligned}
s^2(\theta_0^*) &= \sigma^2 \frac{9m^2 + 9m - 3}{8m^3 + 12m^2 - 2m - 3} = \sigma^2 s_0^2(m), \\
s^2(\theta_1^*) &= \sigma^2 \frac{3}{2m^3 + 3m^2 + m} = \sigma^2 s_1^2(m), \\
s^2(\theta_2^*) &= \sigma^2 \frac{45}{m(2m+1)(4m^3 + 8m^2 + m - 3)} = \sigma^2 s_2^2(m).
\end{aligned} \tag{3.1.5}$$

Note the very strong dependence of the dispersions on the number of points $n = (2m+1)$ used for fitting. At reasonably large n , $s^2(\theta_1^*) \propto n^{-3}$ and $s^2(\theta_2^*) \propto n^{-5}$. Consequently, the relative position error $\delta x_{\min}/|x_{\min}| \equiv \sqrt{s^2(x_{\min})}/|x_{\min}|$ strongly depends on n . In order to illustrate the dependence, the corresponding calculations were performed assuming the point set (3.1.1) such that $x_{\min} \approx 1$ pixel. Then, $|\theta_1^*| \approx 2|\theta_2^*|$ and the normalized relative error of positioning is:

$$\left(\delta x_{\min}/|x_{\min}|\right) \cdot \left(|\theta_2^*|/\sigma\right) = \left(\sqrt{s^2(x_{\min})}/|x_{\min}|\right) \cdot \left(|\theta_2^*|/\sigma\right) \approx \sqrt{s_1^2(m)/4 + s_2^2(m)}. \tag{3.1.6}$$

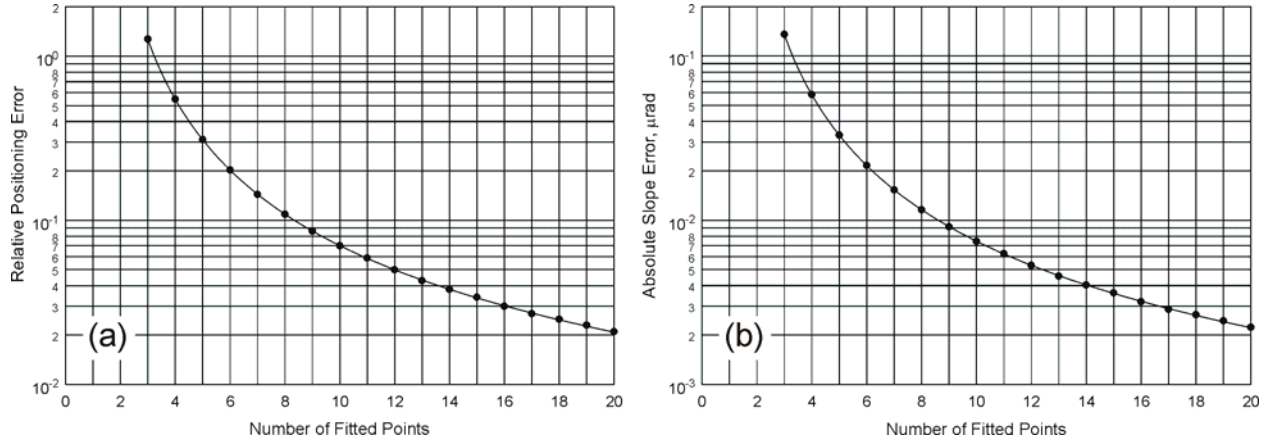


Figure 2. (a) Dependence of the normalized relative error of positioning given by Eq. (3.1.6) on number of points n used for fitting and (b) corresponding absolute error of the LTP slope measurement due to the PDA dark current calculated at the parameters of experiment shown in Fig. 1 (see text for more details).

The result of the calculation is presented in Fig. 2. The noticeable feature is very steep dependence at smaller n . At increasing number of points from $n = 7$ to $n = 15$, the considered random noise is suppressed by factor more than four.

Figure 3 shows the dark signal measured in the LTP sample channel without incident light. The signal is due to the dark current of the PDA. The temporal variation of the dark current is one of the sources for random noise while the light intensity is measured. Variation of the dark current from pixel to pixel is a source for a systematic error. In our case, the random dark current variations seen in Fig. 3a with standard deviation of $\sigma_{dc} \approx 6.8$ bits (ADC digits) are significantly larger than the systematic pixel-to-pixel variations remaining in Fig. 3b after averaging over 34 consequent exposures similar to one shown in Fig. 3a.

Let us evaluate the absolute value of the positioning errors due to the random fluctuations of dark current (Fig. 3a). From the fit parameters presented in Fig. 1, we can calculate the regression parameters $\theta_2^* \approx 639$ bits/(pixel) 2 and $\theta_1^* \approx 2\theta_2^* \cdot 1\text{pixel} \approx 1278$ bits/pixel for the set of points of the kind (3.1.1) shifted to get $x_{\min} \approx 1$ pixel. For $n = 11$ points used to fit the central part of the pattern in Fig. 1, the positioning error due to random dark current noise is

$$\delta x_{\min} \approx x_{\min} \left(\sigma_{dc} / \theta_2^* \right) \sqrt{s_1^2(m)/4 + s_2^2(m)} \approx 6.3 \cdot 10^{-4} \text{ pixels} \approx 1.6 \cdot 10^{-2} \mu\text{m} \tag{3.1.7}$$

that corresponds to very small contribution of $\delta\alpha_{dc} \approx 0.006 \mu\text{rad}$ into the slope error – compare with Fig. 2b.

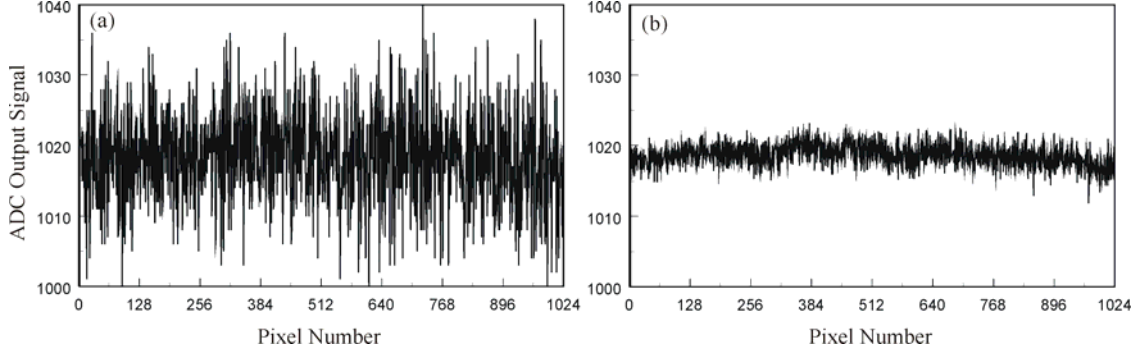


Figure 2: Dark signal in the LTP sample channel: (a) dark signal for one particular exposure; (b) dark signal averaged over 34 consequent exposures. The systematic pixel-to-pixel variations of the dark signal are clearly seen in plot (b). The intensity of light is shown in units of 16-bit analog-to-digit converter (ADC) output signal.

Other sources of the LTP positioning error, e.g. read-out error, electronics noise, which are independent of the value of the light intensity detected, can be treated in the same way.

The read-out error is due to internal switching in PDA and depends on the array design and read-out method. For a PDA similar to one used in our LTP,²² the peak-to-peak fixed pattern (systematic) dark signal due to all switching transient effects is usually stated to be less than 1% of the saturated amplitude.²³ Assuming the saturated amplitude of $6 \cdot 10^4$ ADC output digits (ADC with 10^{16} bits) and a value of the read-out effect of approximately 0.3%, the read-out error can be estimated to be $\sigma_{ro} \approx 200$ ADC digits. The corresponding positioning error at $n = 11$ is [compare with (3.1.7)]

$$\delta x_{\min} \approx x_{\min} \left(\sigma_{ro} / \theta_2^* \right) \sqrt{s_1^2(m)/4 + s_2^2(m)} \approx 2 \cdot 10^{-2} \text{ pixels} \approx 0.5 \mu\text{m}. \quad (3.1.8)$$

The error in slope measurement due to the read-out error estimated in this way is

$$\delta \alpha_{ro} \approx 0.2 \mu\text{rad} \quad \text{at } n = 11. \quad (3.1.9)$$

At the same time, the slope error due to the 2^{16} -bit ADC resolution (~ 1 digit) is only $\delta \alpha_{ADC-16} \approx 0.001 \mu\text{rad}$ ($n = 11$). A detector with significantly lower resolution, say with 2^{10} -bit, or even 2^8 -bit, ADC, can be more than adequate for the LTP application, providing the slope error due to digitization is equal to

$$\delta \alpha_{ADC-10} \approx 0.06 \mu\text{rad} \quad \text{at } n = 11. \quad (3.1.10)$$

The effect of substitution of the LTP detector PDA by another one with pixel pitch smaller by factor of k can be easily understood, if a measurement with such a PDA is thought of as a series of k independent measurements with the original PDA with the smaller pixel size. It is obvious that averaging over k independent measurements provides improvement of positioning error by a factor of \sqrt{k} .

3.2. Positioning error due to photoresponse non-uniformity of the photo-diode array

One of the major LTP systematic effects relates to the PDA photoresponse non-uniformity, that is the relative output signal difference of the pixels under same illumination conditions. The photoresponse non-uniformity can be thought as a pseudo-random distribution of the pixel sensitivity characterized with the expectation value of one and the standard deviation σ_{prn} . As a magnitude of σ_{prn} , one can use a value of non-uniformity of the PDA photoresponse for the neighbor pixels that is usually specified to be about 2-3%.^{22,23}

At the assumption, the regression equation (4) corrected to include the error due to the photoresponse non-uniformity is:

$$\begin{aligned} y_i &= (1 + \xi_i) (\theta_0 + \theta_1 x_i + \theta_2 x_i^2) = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \varepsilon_i, \\ \varepsilon_i &= \xi_i (\theta_0 + \theta_1 x_i + \theta_2 x_i^2) \approx \xi_i (\theta_0^* + \theta_1^* x_i + \theta_2^* x_i^2) \equiv \xi_i y_i^*, \end{aligned} \quad (3.2.1)$$

where y_i^* is the PDA signal determined from the evaluated parameters θ_0^* , θ_1^* , and θ_2^* . Recall that a star label is used to separate a variable from its evaluation. The corresponding dispersion of ε_i is

$$D(\varepsilon_i) \approx D(\xi_i) (y_i^*)^2 = \sigma_{pm}^2 (y_i^*)^2. \quad (3.2.2)$$

The regression (3.3.1) can be resolved by introducing a weight matrix \hat{W}_{pm} with nonzero elements,

$$\omega_{i,i} = (y_i^*)^2. \quad (3.2.3)$$

And the solution of the regression can be found by iteration. First, replacing the y_i^* values in the weight matrix with the measured variables y_i , the parameters θ_0^* , θ_1^* and θ_2^* are evaluated from equation (24). Next, the found values of the polynomial parameters are used to estimate y_i^* and construct the corrected weight matrix, which is used then to find the second approximation for θ_r^* via Eq. (24), and so on. The parameters found during the iteration procedure can be used to evaluate the positioning error via Eqs. (25) and (3.3). The corresponding numerical calculation accounting for $\sigma_{pm} \approx 0.02$ for the interference pattern shown in Fig. 1 gives

$$\delta x_{\min} \approx 0.03 \text{ pixels}; \quad \delta \alpha \approx 0.3 \mu\text{rad} \text{ at } n = 11 \text{ and } \delta x_{\min} \approx 0.02 \text{ pixels}; \quad \delta \alpha \approx 0.2 \mu\text{rad} \text{ at } n = 23. \quad (3.2.4)$$

The dependence of the positioning error on the number of points used for fitting is shown in Fig. 4. The dependence has asymptotic behavior $\propto n^{-1/2}$. Note that the systematic error due to photoresponse non-uniformity of the detector considered in this section can be suppressed if an appropriate calibration procedure of the detector with a flat-field light source is used.

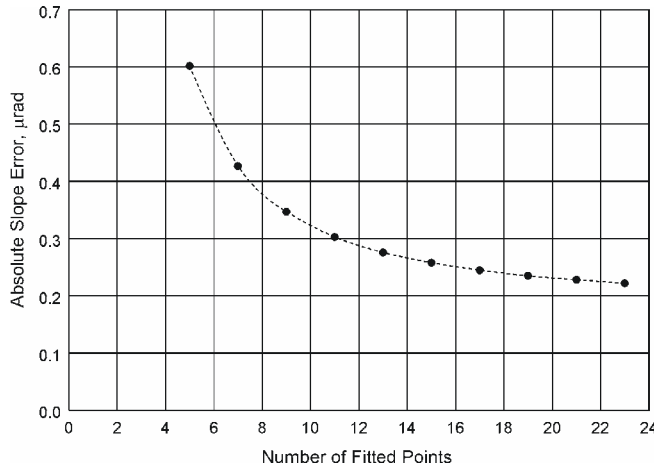


Figure 4: Dependence of the absolute error of the LTP slope measurement due to the photoresponse non-uniformity of the PDA on number of points used for fitting. The calculation was performed for the parameters of experiment shown in Fig. 1.

3.3. Non-uniformity of the effective pixel pitch

Above, we have supposed the exact knowledge of the position of each pixel x_i . However, this assumption can be broken due to the limited tolerance of the detector chip production process. For the used PDA, the possible range for the standard deviation σ_p of the pixel positions from an ideal pattern with uniform pitch distribution can be estimated to be between 0.1 μm and 2 μm . The lower limit comes from the size tolerance characteristic for the CMOS process of the PDA production. The upper limit can be obtained assuming that the photoresponse non-uniformity of $\sim 2\%$ originates from the non-uniformity of the pixel pitch and size.

The non-uniformity of the effective pixel pitch leads to a systematic error of beam positioning. The magnitude of the error can be reasonably estimated if one assumes a pseudo-random character of the pitch non-uniformity. With this assumption, the pixel number variable x_i is a fixed quantity, while the unobserved true value of the i_{th} – pixel position is a random variable. Such a situation, termed the Berkson case (see, e.g.,¹⁸ Chapter 5.7), can be resolved with the standard regression procedure with the error terms in equation (4) corrected to include the error of the pixel position:

$$\begin{aligned} y_i &= \theta_0 + \theta_1 (x_i + \delta_i) + \theta_2 (x_i + \delta_i)^2 + \beta_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \varepsilon_i, \\ \varepsilon_i &= \beta_i + \theta_1 \delta_i + \theta_2 \delta_i^2 + 2\theta_2 \delta_i x_i. \end{aligned} \quad (3.3.1)$$

In the expression for combined error ε_i (3.3.1), the term $\theta_2 \delta_i^2$ can be omitted because of $\delta_i \ll 1$ (according to the limit of $\delta_i \leq 0.1$ pixel mentioned above)

$$\varepsilon_i \approx \beta_i + \theta_1 \delta_i + 2\theta_2 \delta_i x_i = \varepsilon_i + \delta_i (\theta_1 + 2\theta_2 x_i). \quad (3.3.2)$$

Due to independence of β_i and δ_i , the dispersion of ε_i is

$$D(\varepsilon_i) = D(\beta_i) + D(\delta_i)(\theta_1 + 2\theta_2 x_i)^2 \approx D(\beta_i) + \sigma_p^2 (\theta_1^* + 2\theta_2^* x_i)^2. \quad (3.3.3)$$

In the last equation, the dispersion is expressed in terms of evaluated parameters θ_1^* and θ_2^* ; and a pseudo-random distribution of the pixel pitch is accounted with the parameter of pixel position dispersion σ_p^2 . Similar to the case of photoresponse non-uniformity (Sec. 3.2), the solution can be found by iteration. Assuming the same weight for all points, the estimate for θ_1^* and θ_2^* is first evaluated from equation (12). Next, using the found values of the polynomial parameters, the first approximation for the weight matrix is determined, which is used then to find the second approximation for θ_1^* and θ_2^* via Eq. (24), and so on.

In order to estimate a contribution of the pixel pitch non-uniformity into the positioning error, let us assume $D(\beta_i) \equiv 0$. The residual dispersion matrix can then be written as a product of the pixel dispersion parameter σ_p^2 and a diagonal weight matrix \hat{W}_p with nonzero elements

$$\omega_{i,i} = (\theta_1^* + 2\theta_2^* x_i)^2 = (2\theta_2^*)^2 (x_i - x_{\min}^*)^2. \quad (3.3.4)$$

Let us choose the set of variables (x_i) as a centered set of $n = 2(m+1)$ pixels, analogous to (3.1.1), $(x_i) = (-m, -(m-1), \dots, -i, \dots, -1, 0, 1, \dots, i, \dots, (m-1), m)$, and such as $x_{\min} \approx 1/2$ pixel. With such a choice, the matrix $(\hat{A}' \hat{W}_p^{-1} \hat{A})$ has shape similar to (3.1.2)

$$(\hat{A}' \hat{W}_p^{-1} \hat{A}) = (2\theta_2^*)^{-2} \begin{bmatrix} Z0 & 0 & Z2 \\ 0 & Z2 & 0 \\ Z2 & 0 & Z4 \end{bmatrix}, \quad (3.3.5)$$

where

$$Z0 = \sum_{i=-m}^m \frac{x_i^0}{(x_i - x_{\min})^2}; \quad Z2 = \sum_{i=-m}^m \frac{x_i^2}{(x_i - x_{\min})^2}; \quad Z4 = \sum_{i=-m}^m \frac{x_i^4}{(x_i - x_{\min})^2}. \quad (3.3.6)$$

The dispersion matrix of the regression parameters is

$$D(\hat{\theta}) = \sigma_p^2 (\hat{A}' \hat{W}_p^{-1} \hat{A})^{-1} = \sigma_p^2 (2\theta_2^*)^2 \begin{bmatrix} Z0 & 0 & Z2 \\ 0 & Z2 & 0 \\ Z2 & 0 & Z4 \end{bmatrix}^{-1} = \sigma_p^2 (2\theta_2^*)^2 \hat{Z}, \quad (3.3.7)$$

where \hat{Z} is 3×3 matrix with the elements $z_{k,l}(n)$ dependent only on the number of points n used for fitting.

Note $s^2(\theta_1^*) = \sigma_p^2 (2\theta_2^*)^2 z_{2,2}(n)$ and $s^2(\theta_2^*) = \sigma_p^2 (2\theta_2^*)^2 z_{3,3}(n)$. Therefore, the positioning error is

$$\delta x_{\min} / |x_{\min}| \approx 2 \sigma_p \sqrt{z_{2,2}(n) + z_{3,3}(n)} . \quad (3.3.8)$$

Deriving Eq. (3.3.8), we used equality $|\theta_1^*| \approx |\theta_2^*|$, which is valid at $x_{\min} \approx 1/2$ [compare with (3.2)].

For the example of the LTP interference pattern shown in Fig. 1, numerical evaluation gives:

$$z_{2,2}(11) \approx 7.58 \cdot 10^{-2}, \quad z_{3,3}(11) \approx 9.96 \cdot 10^{-3} \quad \text{and} \quad z_{2,2}(23) \approx 3.95 \cdot 10^{-2}, \quad z_{3,3}(23) \approx 1.04 \cdot 10^{-3}. \quad (3.3.9)$$

And at $\sigma_p \approx 2 \mu\text{m}$, $x_{\min} \approx 1/2$

$$\delta x_{\min} \approx 0.73 \mu\text{m}; \delta \alpha \approx 0.3 \mu\text{rad} \text{ at } n = 11 \text{ and } \delta x_{\min} \approx 0.50 \mu\text{m}; \delta \alpha \approx 0.2 \mu\text{rad} \text{ at } n = 23, \quad (3.3.10)$$

suggesting scaling $\propto n^{-1/2}$ at larger n . Recall that the effect of the non-uniformity of the pixel pitch estimated in this section has systematic character and, therefore, the corresponding error cannot be suppressed by multiple measurements. One of the way to suppress the error is to precisely calibrate the pixel pitch with accuracy of $\sim 0.1 \mu\text{m}$. More practical way is to increase the number of the fitted points by replacing the existing PDA with a detector with higher spatial resolution (smaller pixel pitch).

3.4. Positioning error due to signal shot noise

To complete the analysis of the LTP error sources, let us consider the case of signal shot (Poisson) noise:

$$y_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \zeta_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \varepsilon_i k \sqrt{y_i^*}; \quad \varepsilon_i \approx \zeta_i / \sqrt{y_i^*}. \quad (3.4.1)$$

The transformations in (3.4.1) were performed to construct a regression, which can be solved with the Least Square Method. For this purpose, new error variables ε_i with equal variance were built. The variance ε_i is unbiased and independent on the measured signal magnitude. The dispersion of ε_i is $D(\varepsilon_i) = 1$. The constant k is introduced to normalize the shot noise of the photo-electrons to the units of bits of the ADS output signal. In our case, $k \approx 0.02$. The regression (3.4.1) has the normal ‘canonical’ shape for measurement with weighted points described with a diagonal weight matrix \hat{W}_{sn} with nonzero elements given by

$$\omega_{i,i} = y_i^*. \quad (3.4.2)$$

The parameters of the regression (3.4.1) can be found with an iteration procedure. The found parameters can be used to evaluate the positioning error via Eqs. (25) and (3.3). The corresponding numerical calculation for the interference pattern shown in Fig. 1.1 gives

$$\begin{aligned} \delta x_{\min} &\approx 2 \cdot 10^{-4} \text{ pixels}; \quad \delta \alpha \approx 2 \cdot 10^{-3} \mu\text{rad} && \text{at } n = 11, \\ \delta x_{\min} &\approx 1.12 \cdot 10^{-4} \text{ pixels}; \quad \delta \alpha \approx 1 \cdot 10^{-3} \mu\text{rad} && \text{at } n = 23. \end{aligned} \quad (3.4.3)$$

Note that in the considering case of the signal shot noise, the dependence of the positioning error on the number of points used for fitting has asymptotic behavior $\propto n^{3/4}$. This asymptote is due to the fact that at our choice of the symmetrical set of points $\{x_i\}$ (3.1.1), an increase of n correlates with adding the variables y_i with larger values and, therefore, larger statistical weight of the measurement.

4. CONCLUSION ABOUT POSITIONING ERROR AND FITTING STRATEGY

Table 1 summarizes the results obtained for positioning error due to the different sources of LTP II noise. The table also contains the found asymptotic dependence of the effects on the number of points used to determine the position of the interference pattern minimum.

Table 1. LTP II positioning error due to the different sources of noise.

Noise source	Position error (pixel) at $n = 11$	Slope error (μrad)		Asymptotic with n
		$n = 11$	$n = 23$	
PDA dark signal	$6 \cdot 10^{-4}$	0.006	0.002	$\propto n^{-1/2}$
Read-out noise	$2 \cdot 10^{-2}$	0.2	0.07	$\propto n^{-1/2}$
ADC resolution (at 2^{10} bits)	$2 \cdot 10^{-3}$	0.06	0.02	$\propto n^{-1/2}$
PDA photo-response (at 2%)	$3 \cdot 10^{-2}$	0.3	0.2	$\propto n^{-1/2}$
Pixel pitch non-uniformity (at 10%)	$3 \cdot 10^{-2}$	0.3	0.2	$\propto n^{-1/2}$
Signal shot noise	$2 \cdot 10^{-4}$	0.002	0.001	$\propto n^{-3/4}$

A significant improvement of the positioning error can be obtained with a more sophisticated fitting procedure, which has to allow extending the number of points used for fitting. Both the random and the systematic errors have strong dependence on the number of points. Although the dependence has a shape $\propto n^{-1/2}$ asymptotically at large n , in the range of n practically available at the present configuration of the LTP, the dependence is even much stronger.

Replacement of the existing PDA with another one with significantly smaller pixel pitch has an additional advantage. The estimation presented in Sec. 3.1 shows that a PDA camera with 2^{10} -bit ADC is adequate for the LTP measurement with accuracy of $\sim 0.1 \mu\text{rad}$.

Note that in the present work, we have estimated the slope error for one LTP arm. However, the same error sources should be accounted for both the sample and the reference arms. Assuming independence of the errors in the arms, the estimate for the total slope measurement error has to be increased by a factor of $\sqrt{2}$ compared with the values given in Table 1.

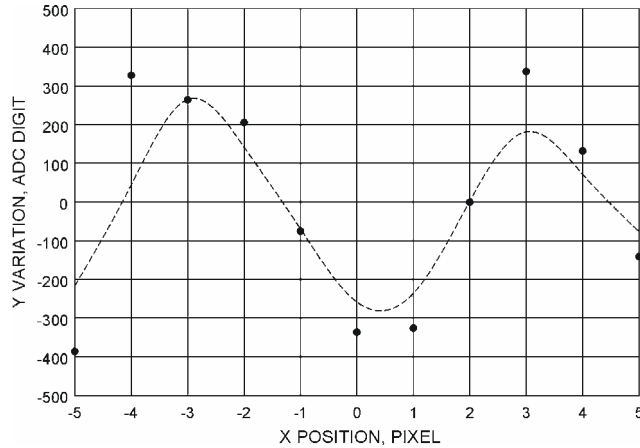


Figure 5: The distribution of the intensity signal variation around the best fit second polynomial. The dashed curve is plotted just to guide the eye (see discussion in the text).

In conclusion, let us discuss the compatibility and reliability of the second-order polynomial to fit the LTP interference pattern. Figure 5 shows the distribution of the intensity signal variations around the best fit second polynomial found by the regression analysis method. The corresponding standard deviation is $\sigma \approx 273$ ADC digits. The dashed curve is plotted just to guide the eye. It seems that the curve suggests the systematical character of the variation distribution. However, the total amount of points, $n = 11$, is too small in order to make a positive conclusion about the systematic character of the variation. An increase of n , desirable for improving the positioning accuracy, can lead to a significant systematic deviation of the second-order polynomial model from the shape of the interference feature under consideration. In this case, a more complicated regression model has to be used.

ACKNOWLEDGEMENTS

The author is grateful to E. Church, S. Irick, A. MacDowell, W. McKinney, H. Padmore, P. Takacs, and T. Warwick for extremely useful discussions. This work was supported by the U. S. Department of Energy under contract number DE-AC02-05CH11231.

REFERENCES

1. E. L. Church, P. Z. Takacs, *Use of an optical profiling instrument for the measurement of the figure and finish of optical quality surfaces*, Wear, 109 (1986), 241-57.
2. P. Z. Takacs, SHINAN Qian, J. Colbert, *Design of a long trace surface profiler*, Proceedings of SPIE 749 (1987), 59-64.
3. P. Z. Takacs, S. K. Feng, E. L. Church, SHINAN Qian, W-M. Liu, *Long trace profile measurements on cylindrical aspheres*, Proceedings of SPIE , 966 (1989), 354-64.
4. S. C. Irick, W. R. McKinney, *Advancements in one-dimensional profiling with a long trace profiler*, Proceedings of SPIE, 1720 (1992), 162-8.
5. S. C. Irick, W. R. McKinney, D. L. Lunt, P. Z. Takacs, *Using a straightness reference in obtaining more accurate surface profiles from a long trace profiler (for synchrotron optics)*, Rev. Sci. Instrum., 63(1) , 1436-8 (1992).
6. S. C. Irick, *Improved measurement accuracy in a long trace profiler: compensation for laser pointing instability*, Nuclear Instruments & Methods in Physics Research A-Accelerators Spectrometers Detectors & Associated Equipment, 347(1-3), 226-30 (1994).
7. S. C. Irick, *Determine surface profile from sequential interference patterns from a long tracer profiler (for synchrotron optics)*, Rev. Sci. Instrum. 63(1) (1992) 1432-5.
8. SHINAN Qian, H. Li, P. Z. Takacs, *Penta-Prism Long Trace Profiler (PPLTP) for measurement of grazing incidence space optics*, Proceedings of SPIE, 2805 (1996), 108-14.
9. P. Z. Takacs, SHINAN Qian, T. Kester, H. Li, *Large-mirror figure measurement by optical profilometry techniques*, Proceedings of SPIE, 3782 (1999), 266-74.
10. SHINAN Qian, G. Sostero, P. Z. Takacs, *Precision calibration and systematic error reduction in the Long Trace Profiler*, Proceedings of SPIE, 3782 (1999), 627-36.
11. Shinan Qian, Sostero G, P. Z. Takacs, *Precision calibration and systematic error reduction in the long trace profiler*, Optical Engineering, 39(1) (2000), 304-10.
12. P. Z. Takacs, SHINAN Qian, *Accuracy limitations in long-trace profilometry*, AIP Conference Proceedings, 708 (2004), 831-4.
13. SHINAN Qian, W. Jark, P. Z. Takacs, *The penta-prism LTP: a long-trace-profiler with stationary optical head and moving penta prism*, Rev. Sci. Instrum. 66(3) (1995), 2562-9.
14. P. Z. Takacs, C. J. Bresloff, *Significant improvements in long trace profiler measurement performance*, Proceedings of SPIE 2856 (1996), 236-45.
15. P. Z. Takacs, E. L. Church, C. J. Bresloff, L. Assoufid, *Improvements in the accuracy and the repeatability of long trace profiler measurements*, Applied Optics, 38(25) (1999), 5468-79.
16. Shinan Qian, P. Z. Takacs, *Equal optical path beamsplitter for a pencil beam interferometer and shearing interferometer*, Optical Engineering 42(no.4) (2003).929-34.
17. V. V. Yashchuk, S. C. Irick, A. A. MacDowell, *Elimination of 'ghost'-effect-related systematic errors in metrology of X-ray optics with a long trace profiler*, Proceedings of SPIE 5858-34 (2005).
18. R. L. Plackett, *Principles of Regression Analysis* (Oxford, At The Clarendon Press, 1960).
19. D. J. Hudson, *Statistics: Lectures on Elementary Statistics and Probability* (Geneva, 1964; In Russian: D. J. Hudson, *Statistics for Physicists*, Moscow, Mir, 1970).
20. J. Neter and W. Wasserman, *Applied Linear Statistical Models* (London, Inwin-Dorsey International, 1974).
21. M. Kendall and A. Stuart, *The Advance theory of Statistics*, vol.2 (New York, Oxford University Press, 1979).
22. Hamamatsu Photonics K.K., *Specification: NMOS Linear Image Sensor S3901/S3904 series* (<http://usa.hamamatsu.com>).
23. PerkinElmer Optoelectronics Inc., *RETICONTM; L-series Linear CMOC Spectroscopy Sensor* (www.perkinelmer.com/opto).